APPENDIX B
REPLACEMENT PAGES

exon and regulatory sequence of genomic nucleic acid, and distinctions between conventional genomic and mRNA transcript sequence, and viral genomic transcript and reverse transcriptase transcript.

Thus for the analyte sequence (the ribonucleotide U and the deoxyribonucleotide T are used interchangeably for base pairing purposes) 5'-ATAAAGCTGCTTC (SEQ ID. NO. 1) (having no subfragments) will hybridize only to beads or array sites having the 5-mers

5'-ATAAA , 5'-TAAAG , 5'-AAAGC ,

5'-AAGCT , 5'-AGCTG , 5'-GCTGC , 5'-CTGCT , 5'-TGCTT , and 5'-GCTTC under stringency conditions permitting no mismatch among the five nucleotides available for base pairing. There are 4^5 or 1024 possible 5-mers that can be arrayed on a substrate or present attached to individual beads, but even those similar to the nine perfectly matching 5mers listed above will have sufficiently different energies of hybridization that under stringent conditions analysis of the hybridization data directly will permit sequencing the analyte nucleic acid sequence. Much longer unknown sequences can be readily sequenced segment by segment in this manner, with appropriate consideration of the subfragment problem. In some cases the subfragment ordering may require application of another sequencing method, such as the ligation signature hybridization method (below) and traditional gel electrophoresis methods (Maxam and Gilbert (1977) *supra*; Sanger, et al. (1977) *supra*).

Another sequencing method that relies upon hybridization employs a label or tag that identifies the specific hybridizing sequence. For example a different fluorescent marker can linked to each possible sequence of three nucleotides (4^3 or 64 in all), and a sequence may be obtained by successive hybridization and digestion three nucleotides at a time. The sequence may also be obtained by labels comprising a nucleotide sequence, for example the start codon AUG may be labeled by the sequence 5'-AAAAAAAAACCCCCTTTTCTTTT (SEQ ID NO: 2), which will form a hairpin loop self complementary structure that can be differentiated from like labeling structures, such as 5'-AAAAAAAAACCCCCTTTTTTTTT) (SEQ ID NO: 3) and 5'-AAAAGAAAACCCCCTTTTCTTTT (SEQ ID NO: 4), by the temperature that causes a loss of such secondary structure.

base pairing properties permit hybridization between a sequence containing the 8-oxo-dG at a position in the sequence and an A in the corresponding position. Although the base degenerate base pairing properties of the deoxyribonucleoside triphosphate analogs 8-oxo-dG and dPTP are employed in an indeterminate or non-determinate manner to induce the random mutagenesis described above, nucleotides comprising nucleosides having degenerate complementarity sets that partially overlap as do the base pairing complementarity sets of 8-oxo-dG (base pairing complementarity set = {C, A}) and dPTP (base pairing complementarity set = {G, A}), which overlap in the common A and both exclude the nucleotide T, can be used in a determinate manner.

Likewise, a specific sequence position of two probes, each having partially overlapping base pairing sets of two possible nucleotides at that sequence position, such as two probes for hybridization having a sequence 5'-AT(X_i)GG linked to a chemiluminescent (ChL) or other tag, 5'-AT(X_1)GG-CL_1 and 5'-AT(X_2)GG-CL_2, where ChL_1 and ChL_2 are chemiluminescent at different frequencies, and X_1 comprises T or C in equal proportions, and X_2 comprises G or T in equal proportions making the third (X_1) position of 5'-AT(X_1)GG-ChL_1 pair degenerately to the set of nucleotides G and A (base pairing complementarity set = {G, A}), and the third (X_2) position of 5'-AT(X_2)GG-ChL_2 pair degenerately to the set of nucleotides C and A (base pairing complementarity set = {C, A}). Thus 5'-AT(X_2)GG-ChL_2 is the effective equivalent to the degenerately pairing hybridization probe 5'-AT(dP)GG-ChL_2, which utilizes, instead of equal proportions at the third position of C and T, the deoxynucleoside analog dP which base pairs, for the purpose of hybridization, almost equally with G and A. Analogously 5'-AT(X_1)GG-ChL_1 is the equivalent to 5'-AT(8-oxo-dG)GG-ChL_1, with the degenerately pairing analog 8-oxo-dG, which pairs, for the purposes of hybridization, nearly equally with A and C, at the third position instead of equal proportions of T and G. Both sets of hybridization probes {5'-AT(dP)GG-ChL_2, 5'-AT(8-oxo-dG)GG-ChL_1} and {5'-AT(X_2)GG-ChL_2, 5'-AT(X_1)GG-ChL_1} as well as sets in which a degenerately base pairing nucleoside analog is employed for one of the probes, while equal proportions of nucleosides having the desired base pairing properties may be

employed, as long as the base pairing sets overlap in the manner described, e.g for two doubly degenerate base pairing sets, overlap of one of the nucleotides. Two unique doubly degenerate base pairing sets, e.g. each base pairing complementary set containing two nucleosides that are about equally paired for hybridization purposes, are required for normal nucleic acid sequences having four possible nucleotides (the ribonucleoside Uracil (U) being equivalent for these purposes to T).

If the sequence to be analyzed contains or may contain additional nucleotides, more sets having overlap are required for determinate use of the degenerate base pairing. For example, if six nucleotides could be in the sequence, five quadruply degenerate pairing probes could be employed. Each of these five probes must have at the position of interest or probed position a unique base pairing set containing one of the six possible nucleotides, so that all the sets contain the specific nucleotide, and one of the six possible nucleotides must be absent from all the base pairing sets. Further, each unique base pairing set, in addition to overlapping with the remaining four base pairing sets in the nucleotide common to all five sets, for example, also overlaps in two other of the possible nucleotides with any other probe. This additional overlap of two nucleotides cannot be the same for all pairs of quadruply degenerate probes if all the base pairing sets are unique. In this manner all five quadruply degenerate pairing probes would hybridize to the specific sequence in which the common base of the base pairing set is present at the position of interest, and none of the probes would, under appropriately stringent hybridization conditions, hybridize to the sequence in which the base absent from all five quadruply degenerate base pairing sets is present at the position of interest. When the other four nucleotides are present at the position of interest, the system is constructed such that four of the five specific probes will hybridize to the analyte sequence.

The situation is much simpler for the typical case of four possible nucleotides in a hybridizing sequence, where two probes having unique doubly degenerate partially overlapping base pairing sets at one position may be employed in a determinate fashion. For example probe sets such as {5'-AT(dP)GG-ChL_2 , 5'-AT(8-oxo-dG)GG-ChL_1 }, {5'-AT(X_2)GG-ChL_2 , 5'-AT(X_1)GG-ChL_1}, {5'-AT(X_2)GG-ChL_2 , 5'-AT(8-oxo-dG)GG-ChL_1} and {5'-AT(dP)GG-ChL_2, 5'-AT(X_1)GG-ChL_1} could be used to probe for the antiparallel sequence 5'-CCξAT where ξ is an unknown or variable base at the sequence

position of interest or variable position. If ξ is T, none of the probes will hybridize to the analyte sequence, while both probes will hybridize to the analyte if ξ is A. If the identity of ξ is G only one of the probes will hybridize (either 5'-AT(dP)GG-ChL_2 or 5'-$AT(X_2)$GG-ChL_2 depending upon which is employed), and if ξ is C only the other (only one) of the two probes will hybridize (either 5'-AT(8-oxo-dG)GG-ChL_1 or 5'-$AT(X_1)$GG-ChL_1 again depending upon which is employed). This permits use of two probes instead of four if non-degenerate probes were employed with full knowledge of the identity of ξ and therefore a determinate use of the degenerate probes.

The preceding has been described in the context of tagged or labeled hybridization probes which may be employed for sequencing using tagged probes. First that the label or tag need not be chemiluminescent should be noted. For example a fluorescent or otherwise spectroscopically detectible tagging moiety may be employed. Alternatively the sequence that is expected to hybridize may be tagged or labeled with a nucleic acid sequence that does not hybridize by virtue of its properties, for example the tendency to form hairpin loops or some other non-hybridizing structure or a sequence that is known not to be complementary to any sequence in the analyte, such as polyA or polyT for genomic analyte (where mRNA tails are not present). Further, two "colors" or spectroscopically detectible frequencies of chemiluminescense are also described above, and facilitate a two color assay akin to two color hybridization as described in U.S. Patent No. 5,800,992 to Fodor et al. Although employing two colors facilitates probing simultaneously with the two probes by permitting simultaneous visualization of the two probes rather than multiple detection steps, to detect analyte sequences hybridizing to one (1^{st} frequency) the second (2^{nd} frequency) or both (composite of the two frequencies) probes, this is not requisite for practicing the invention. The two probes may be employed sequentially with a conventional tagging or labeling moiety that is the same for both probes. Additionally the probes need not be tagged or labeled by a discrete labeling moiety as is the case when methods for sequencing by hybridization that do not employ discrete tags or labels are employed (U.S. Patent No. 5,525,464 to Drmanac et al.), and hybridization may be detected by detecting ^{32}P autoradiographically. Alternatively hybridization can be detected without any label, whether a separate moiety or part of the nucleic acid, even the incorporation of ^{32}P into probe or analyte, by thermal detection, as when an oligonucleotide array of probes is hybridized to analyte while

For example, the MPSS adapters taught by Brenner et al., 16 adapter sequences having four nucleotide overhangs (overhang position indicated):

(i) adapter position four (analyte "base 1"):

5'-NNNA , 5'-NNNG , 5'-NNNC ,

5'-NNNT ;

(ii) adapter position three (analyte "base 2"):

5'-NNAN , 5'-NNGN , 5'-NNCN ,

5'-NNTN ;

(iii) adapter position two (analyte "base 3"):

5'-NANN , 5'-NGNN , 5'-NCNN ,

5'-NTNN ;

(iv) adapter position one (analyte "base 4"):

5'-ANNN , 5'-GNNN , 5'-CNNN ,

5'-TNNN ,

where N represents any of A or G or C or T(U).

The sixteen adapter sequences listed above are actually adapter sets, each adapter set having 4^3 (64) nucleic acid sequences by virtue of N being any of four nucleotides. These sets can be replaced by eight adapter sequence sets having the sequences listed below. Every four adapter sets corresponding to a specific position of interest or variable position can be replaced by a pair of adapter sets, and each group of four sequences from these four adapter sets that differ only at one position can be replaced by a pair of overhang sequence adapters. Because the MPSS adapters described by Brenner et al., employ ten nucleotide long sequences to tag or label the adapters termed $\mathbf{F_n}$ by the authors, the overhang sequences linked to the $\mathbf{F_n}$ sequences by a common 14 nucleotide long linking nucleic acid sequence 5'-ACGAGCTGCCAGTC-3' (SEQ ID NO: 5).

Because each F_n sequence, which is detected in the MPSS method of Brenner et al. by hybridization to one of the 256 F_n decoder binding site sequences, which number 16 unique sequences, four (signifying the four possible nucleotides) for each overhang position, to the complementary phycoerythrin labeled (PE-labeled) decoder probes, which also number 4 for each position (thus 4*4 or 16 unique sequences *en toto*, and thus 16 adapters, or adapter groups, and PE-labeled decoder probes). For each ligation step of the MPSS method, in which one of the four possible positions is probed or determined, sixteen decoder probes, one for each of the sixteen adapter sequence

groups having the overhang sequences depicted above are hybridized to the decoder binding sites of the encoded adapters in sixteen hybridization cycles, and the arrayed beads are imaged after each such hybridization.

The methods and degenerately base pairing sequences of the instant invention permit halving the number of adapters used and consequently halving the total number of decoder binding sequences and complementary PE-labeled decoder probes, and halving the number of subcycles required to image a ligation cycle and the number of PE-labeled decoder probes per ligation cycle. Using two color labels for decoder probes can reduce the number of subcycles in half again. Additionally possible is the use of partially overlapping unique doubly degenerate sequence positions in the labeled decoder probe sequences to replace four decoder sequences with a pair and further reduce the number of PE-labeled decoder probe sequences directly.

The adapter probe sequences (sequence sets as N is A, T(U), G or C) employed by the method of the instant invention are:

(i) adapter position four (analyte "base 1"):

5'-NNNψ_1 , 5'-NNNψ_2 ;

(ii) adapter position three (analyte "base 2"):

5'-NNψ_1N , 5'-NNψ_2N ;

(iii) adapter position two (analyte "base 3"):

5'-Nψ_1NN , 5'-Nψ_2NN ;

(iv) adapter position one (analyte "base 4"):

5'-ψ_1NNN , 5'-ψ_2NNN .

In the preceding sequences, ψ_1 represents a position having, for example, the doubly degenerate base pairing set {A, G} and ψ_2 position having, for example, the doubly degenerate base pairing set {G, C}. Any of the ψ_1 and ψ_2 doubly degenerate base pairing sets listed in Table 1 below may be employed for ψ_1 and ψ_2.

For example ψ_1 may have the doubly degenerate base pairing set {A, G}, and ψ_2 may have the doubly degenerate base pairing set {A, C}, in which case ψ_1 may be dP and ψ_2 may be 8-oxo-dG. Alternatively, for ψ_1 having the doubly degenerate base pairing set {A, G}, and ψ_2 having the doubly degenerate base pairing set {A, C}, ψ_1 may be $\mathbf{X_1}$ and ψ_2 may be $\mathbf{X_2}$, $\mathbf{X_1}$ being about equal amounts of T and C and $\mathbf{X_2}$ being about

equal amounts of T and G as described above. Or for the same ψ_1 and ψ_2 base pairing sets, ψ_1 may be $\mathbf{X_1}$ and ψ_2 may be 8-oxo-dG, or ψ_1 may be dP and ψ_2 may be $\mathbf{X_2}$. Those of ordinary skill in the art will appreciate that to obtain the same signal intensity from hybridization of $\mathbf{X_1}$ and $\mathbf{X_2}$ type probes as from degenerately pairing nucleotide probes such as those incorporating P or 8-oxo-G, about twice as much probe will be required because only half of the $\mathbf{X_1}$ or $\mathbf{X_2}$ probe can hybridize to a sequence within the probes complementary set, while substantially all of the dP or 8-oxo-G probe can hybridize to analyte sequence in the respective complementary sets.

If ψ_1 has the doubly degenerate base pairing set {A, G}, and ψ_2 has the doubly degenerate base pairing set {G, C}, if both ψ_1 and ψ_2 probes bind, then the analyte nucleic acid sequence position is occupied by G. If the ψ_1 probe binds and ψ_2 probe does not bind, then the identity of the base at probed position is A, while if the ψ_2 probe binds and ψ_1 probe does not bind, the identity of the base at probed position is C. If neither probe hybridizes, the identity of the base at the probed position is T. The process is repeated with ψ_1 and ψ_2 probes for each overhang position (four in all for the instant invention modified MPSS method). If it is desirable to detect a signal for every case, a ninth adaptor, 5'-AAAA, may be included.

Analogously, in the case that ψ_1 has the doubly degenerate base pairing set {A, G}, and ψ_2 has the doubly degenerate base pairing set {A, C}, e.g. ψ_1 is dP and ψ_2 is 8-oxo-dG, if both ψ_1 and ψ_2 probes bind, then the analyte nucleic acid sequence position is occupied by A. If the ψ_1 probe binds and ψ_2 probe does not bind, then the identity of the base at probed position is G, while if the ψ_2 probe binds and ψ_1 probe does not bind, the identity of the base at probed position is C. If neither probe hybridizes, the identity of the base at the probed position is again T. The process is repeated with ψ_1 and ψ_2 probes for each overhang position (four). If it is desirable to detect a signal for every case, a ninth adaptor, 5'-AAAA, may be included.

There are many pairs of partially overlapping doubly degenerate base pairing sets that accomplish substantially the same result. The common element is that they use pairs of hybridization probes that, on the average, hybridize to about 1/2 of the sequences. In the nine hybridization probe case described above, only eight of the probe sets (of 4^3 or 64 sequences each) hybridize to half the beads. The ninth only hybridizes to 1/256. There are more complex code makes all nine probes about equal. To do this, each

Alternatively for transcribed sequence elements, selective expression and cDNA analysis may be used to analyze alleles by the invention. Quantification can be calibrated against known sequence genomic alleles for better calibration of quantification.

As mentioned above, the following 12 pairs of degenerate base pairing sets for ψ_1 and ψ_2 can be employed, in the preceding sequences, or in longer analogous sequences, with the "Ultimate Check Probe" with the indicated base sequence with the complementary base (base not base pairing with either ψ_1 or ψ_2) in parentheses. Additional Check Probes having the sequence 5'-NNZN , where Z base pairs with the base represented in neither $\psi_1$or ψ_2 base pairing set, may be employed for each pair of probes such as (5'-NNψ_1N , 5'-NNψ_2N), to decrease errors further, albeit with an additional probe for each two probes each having ψ_1 or ψ_2 degenerately pairing at one position instead of an additional check probe for the complete set of q pairs of probes for a probed sequence q nucleotides in length. For example q = 4 in the preceding sequences, Z_q is the sequence 5'-ZZZZ, representing the Ultimate Check Probe, which ensures that a sequence not hybridizing to any probes in the set of paired degenerately hybridizing probes {(5'-ψ_1NNN, 5'-ψ_2 NNN), (5'-Nψ_1NN , 5'-Nψ_2NN), (5'-NNψ_1N, 5'-NNψ_2N , (5'-NNNψ_1 , 5'-NNNψ_2)}, is actually a nucleic acid sequence. The set of check probes {5'-ZNNN, 5'-NZNN, 5'-NNZN, 5'-NNNZ} may be employed to check each pair of degenerately hybridizing probes, decreasing error at a cost of an increased number of probes.

and are not intended to limit the scope of what the inventors regard as their invention. Efforts have been made to ensure accuracy with respect to numbers (e.g., amounts, temperature, etc.) but some errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, temperature is in °C and pressure is at or near atmospheric.

In these examples, the following abbreviations have the following meanings:

Å=Angstrom (0.1 nm)
C=Centigrade
kg=kilogram
M=Molar
mg=milligram
ml=milliliter
mm=millimeter
N=Normal
nm=nanometers

Example 1: Preparation of Nucleic Acid Sequences for MPSS Adapters

Oligonucleotides are either purchased presynthesized from Genetic Designs, Inc. Houston, Texas or made on an Applied Biosystems 381A DNA synthesizer. All sequences used are purified by HPLC or gel electrophoresis, which may optionally be omitted.

The following adapter sequences are MPSS encoded adapters of the instant invention for reducing the number of encoded adapters required for the MPSS method. The four nucleotide overhangs are indicated in bold and the decoder binding sequence tag or label is underlined. These are connected by the common sequence 5'-ACGAGCTGCCAGTC (SEQ ID NO: 5), and the common sequence is doble stranded, being hybridized to the complementary sequence 5'-GACTGGCAGCTCGA (SEQ ID NO: 6). The adapter sequences are listed in groups based on their probing and coding for different sequence positions corresponding to the overhang position as in the MPSS method in general, with pairs of adapters having positions with doubly degenerate partially overlapping base pairing sets according to the instant invention instead of the four adapters of MPSS practiced without the instant invention. Thus the adapters include those with doubly degenerate base pairing nucleotides having partially overlapping base pairing sets, and adapters having about equal proportions of two nucleotides at the doubly degenerate base pairing position. The adapters with doubly degenerate base pairing nucleotides having partially overlapping

base pairing sets incorporate dP and 8-oxogG because of their appropriate base pairing properties for the practice of the invention and commercial availability, are organized by probed position as follows:

Overhang position 4, analyte base 1:

5'-**NNN(dP)**ACGAGCTGCCAGTCCATTTAGGCG (SEQ ID NO: 7);

5'-**NNN(8-oxo-dG)**ACGAGCTGCCAGTCCGCTTTGTAG (SEQ ID NO: 8);

Overhang position 3, analyte base 2:

5'-**NN(dP)N**ACGAGCTGCCAGTCGGAACCTGAA (SEQ ID NO: 9);

5'-**NN(8-oxo-dG)N**ACGAGCTGCCAGTCATTCCTCCTC (SEQ ID NO: 10);

Overhang position 2, analyte base 3:

5'-**N(dP)NN**ACGAGCTGCCAGTCCGAAGAAGTC (SEQ ID NO: 11);

5'-**N(8-oxo-dG)NN**ACGAGCTGCCAGTCGGCGATAACT (SEQ ID NO: 12);

Overhang position 1, analyte base 4:

5'-**(dP)NNN**ACGAGCTGCCAGTCGCATCCATCT (SEQ ID NO: 13);

5'-**(8-oxo-dG)NNN**ACGAGCTGCCAGTCGCCAGTGTTA (SEQ ID NO: 14),

where N is A, T(U), G or C.

Also synthesized are the following, grouped by probed position:

Overhang position 4, analyte base 1:

5'-**NNN(X_1)**ACGAGCTGCCAGTCCATTTAGGCG (SEQ ID NO: 15);

5'-**NNN(X_2)**ACGAGCTGCCAGTCCGCTTTGTAG (SEQ ID NO: 16);

Overhang position 3, analyte base 2:

5'-**NN(X_1)N**ACGAGCTGCCAGTCGGAACCTGAA (SEQ ID NO: 17);

5'-**NN(X_2)N**ACGAGCTGCCAGTCATTCCTCCTC (SEQ ID NO: 18);

Overhang position 2, analyte base 3:

5'-**N(X_1)NN**ACGAGCTGCCAGTCCGAAGAAGTC (SEQ ID NO: 19);

5'-**N(X_2)NN**ACGAGCTGCCAGTCGGCGATAACT (SEQ ID NO: 20);

Overhang position 1, analyte base 4:

5'-**(X_1)NNN**ACGAGCTGCCAGTCGCATCCATCT (SEQ ID NO: 21);

5'-**(X_2)NNN**ACGAGCTGCCAGTCGCCAGTGTTA (SEQ ID NO: 22),

where N N is A, T(U), G or C, and X_1 is C or T in equal proportions, and X_2 is G or T in substantially equal proportions.

Using this method (Drmanac et al. U.S. Patent No. 5,695,940) characterization of the thermal stability of short oligonucleotide hybrids was determined on a prototype octamer with 50% GC content, i.e. probe of sequence 5'-TGCTCATG . The theoretical expectation is that this probe is among the less stable octamers, in the 50th percentile or below in stability. Its transition enthalpy is similar to those of more stable heptamers and probes as short as 6 nucleotides in length Bresslauer et al. (1986) *Proc. Natl. Acad. Sci. U.S.A.* 83: 3746. The stability of the 8 bp oligonucleotide duplex hybrid as a function of temperature is evidenced: Parameter $\mathbf{T_d}$, the temperature at which 50% of the hybrid is melted in unit time of a minute is 18 °C. The result shows that $\mathbf{T_d}$ is 15 °C lower for the 8 bp hybrid than for an 11 bp duplex (Wallace et al. (1979) *Nucleic Acids Res.* 6: 3543).

Lane et al. describe a method of measuring thermodynamic parameters of hybridization for nucleic acid probe design in U.S. Patent No. 6,027,884. Absorbance versus temperature profiles (optical melting curves) were collected for each of the molecules at heating and cooling rates of 60 °C per hour over the temperature range from 5 to 85 °C. A data point is collected about every 0.1 °C. Melting curves for samples are collected as a function of total strand concentration, $\mathbf{C_T}$, over a 200 fold range from approximately 500 nM to 100 μM. Absolute absorbance readings ranged from 0.08 OD to 1.3 OD. Optically matched quartz cuvettes with 1 and 0.1 cm path lengths are employed. Such optical nucleic acid melting curves are entirely reversible upon cooling at the same rate. The optical melting curves are normalized to upper and lower baselines and converted to θ_B (the fraction of duplex molecules) versus temperature curves. From these curves the melting or transition temperature, $\mathbf{T_m}$, was determined as the temperature where $\theta_B = \mathbf{0.5}$. These θ_B versus **T** curves may then be analyzed assuming the transitions occur in an "all-or-none" or "two-state" manner, permitting evaluation of the transition by a van't Hoff plot of $\mathbf{1/T_m}$ versus $\mathbf{\ln(C_T)}$. The linear equation describing the resulting plot is:

$$\mathbf{1/T_m = (R/\Delta H)\ln(C_T) + \Delta S/\Delta H} \quad (1).$$

The slope of the van't Hoff plot yields $\mathbf{R/\Delta H}$ and the intercept provides $\mathbf{\Delta S/\Delta H}$. The experimentally determined total free energy is then determined from ΔH and ΔS values at 298.15 °K by $\Delta G_T = \Delta H - T^*\Delta S$ (2).

Thermodynamic parameters of the melting transitions of hybridized nucleic acids

experimentally characterized for $\mathbf{T_m}$ and other thermodynamic parameters of hybridization.

As each probe of the invention generally comprises at the probed position either equal amounts of two more nucleotides, or a degenerately pairing nucleotide analog such as dP, 8-oxo-dG or Inosine (I), which can pair with A, U and C in mRNA-tRNA wobble (G pairing with U and C in wobble) interactions, and is likely similar to 8-oxo-dG, the degenerate fully complementary hybridizations of these probes are expected to have slightly different stabilizations which will affect $\mathbf{T_m}$ to an extent dependent upon the hybridization conditions and oligomer length.

Sequences of the invention having one position corresponding to a probed position are constructed with doubly degenerate base pairing sets given in Table 1 and hybridized to sequences perfectly complementary or mismatched at the probed position to the specific ψ_i doubly degenerate base pairing set. Thus, for example, all pairs of 7-mer probes 5'-NNNψ_1 NNN and 5'-NNNψ_2NNN indicated by the sets of nucleotides ψ_1 and ψ_2 given in Table 1 above are constructed, each 7-mer actually representing a group of 7-mers having 4^6 (4096) sequences (N is one of A T(U) G or C) , and experimentally hybridized under the various conditions.

For the probes of the invention wherein ψ_1 is dP and ψ_2 is 8-oxo-dG, the 4096 7-mer sequences having dP centrally located and the 4096 sequences having 8-oxo-dG centrally located, e.g. at position 4 of 7, correspond respectively to the 5'-NNNψ_1NNN and

5'-NNNψ_2NNN probes. For probes of the instant invention wherein no nucleotide or nucleotide analog (such as dP and 8-oxo-dG) capable of pairing to two nucleotides is incorporated, for example the probes of the invention wherein ψ_1 is $\mathbf{X_1}$ and ψ_2 is $\mathbf{X_2}$, the 4096 7-mer sequences having $\mathbf{X_1}$ centrally located and the 4096 sequences having $\mathbf{X_2}$ centrally located, where $\mathbf{X_1}$ and $\mathbf{X_2}$ are defined as above ($\mathbf{X_1}$ is equal amounts of T and C and $\mathbf{X_2}$ is equal amounts of G and T), correspond respectively to the 5'-NNNψ_2NNN and 5'-NNNψ_1NNN probes. Analogously, probes of the type 5'-Nψ_iNNNNN and ψ_iNNNNNN represent asymmetric internal and terminal probed position probes because the ψ_i position of the probe, corresponding to the probed position, is an asymmetric internal or terminal position respectively.

Each probe using the two possible nucleotides at the probed position is actually a mixture of two hybridizing sequences in about equal proportion, e.g an $\mathbf{X_1}$ probe {1:1}--{5'-GCT(**T**)CAG, 5'-GCT(**C**)CAG} is the equivalent of the single sequence probe incorporating the doubly degenerately pairing nucleotide dP, GCT(**dP**)CAG. Consequently, the stoichiometric equivalent, in terms of hybridization, of the truly doubly degenerate complementarity probes, such as GCT(**dP**)CAG is twice that of probes comprising mixtures having equal nucleic acid content, e.g. 1M of GCT(**dP**)CAG is the stoichiometric equivalent for the purposes of hybridization of the "1M" probe, GCT($\mathbf{X_1}$)CAG, which is actually a mixture of 1M 5'-GCT(**T**)CAG and 1M 5'-GCT(**C**)CAG, and thus 2M in nucleic acid.

For the thermodynamic parameters depending on concentration, stoichiometric equivalents are compared. Each probe is experimentally hybridized to base pair matched sequences at all positions other than the probed position, with the probed hybridizing sequences comprising any of the standard nucleotides A, T(U), G, C. Thus each of the pair of probes 5'-GCT(ψ_i)CAG is hybridized experimentally with:
5'-GCT(**T**)CAG; 5'-GCT(**C**)CAG; and
5'-GCT(**A**)CAG; and 5'-GCT(**C**)CAG. As ψ_i probes that are specified in Table 1 are doubly degenerate, there will be two experimental hybridizations that match at the probed position, these being perfect sequence complementarity, and two hybridizations that mismatch at the probed position, these being single mismatch complementarity. Ideally, a large difference in $\mathbf{T_m}$ will exist between the single mismatch and perfect sequence complementarity experimental hybridizations, representing a large thermodynamic destabilization under the applicable conditions, thus permitting an identifiable distinction to be made between a match and mismatch at the probed position. In addition to varying the conditions to alter the mismatch destabilization magnitude, conditions that affect the total stabilization from hybridization, such as amount of tetramethylammmonium chloride for a GC rich probe, or probe length of the can be varied to decrease or increase total stabilization. Varying the total stabilization, affects the relative amount of the destabilization from the mismatch, reflected in an increased or decreased $\mathbf{T_m}$ depression from the mismatch (corresponding to increased or decreased magnitude of $\Delta\mathbf{T_m}$). Typically, probe lengths

the doubly degenerate base pairing complementarity set: {A, G}. The decoder binding sequences follow: 5'-CATTTAGGCG (SEQ ID NO 23); 5'-GGAACCTGAA (SEQ ID NO: 24); 5'-CGAAGAAGTC (SEQ ID NO: 25); 5'-GCATCCATCT (SEQ ID NO: 26). The corresponding complementary decoder probe sequences (italics) are consequently: 5'-*CGCCTAAATG* (SEQ ID NO: 27); 5'-*TTCAGGTTCC* (SEQ ID NO: 28); 5'-*GACTTCTTCG* (SEQ ID NO: 29); 5'-*AGATGGATGC* (SEQ ID NO: 30).

Oligonucleotide 5'-*CGCCTAAATG* (SEQ ID NO: 27), which has a 5' amino linker, (20 nmoles) in 0.15 ml of water Is treated at room temperature under nitrogen with 0.15 ml of 0.2 M carbonate buffer, pH 8.5 and 0.45 ml of N,N-dimethylformamide (DMF) to give a homogenous solution. To this solution Is added a total of 1.9 mg (3.0 μmoles) of LEAE-NHS in 0.15 ml of DMF in three equal portions, each in a one hour interval. After the addition of the final portion of the LEAE-NHS, the solution was protected from light and stirred at room temperature overnight. The solution was then treated with 2 ml of water and centrifuged at 13,000 RPM for 5 minutes.

The supernatant is passed through a Sepahadex G-25 column (1 x 40cm), eluted with water. The very first peak was collected and concentrated in a rotary evaporator at temperature below 35 °C. The concentrate is separated on a reverse-phase HPLC column (Brownlee, C-8, RP-300, 4.6 x 250 mm), eluted with solvent gradient: 5 to 25% B for15 minutes, followed by 25 to 35% B for 15 minutes, 35 to 60% B for 10 minutes and 60 to 100% B for 5 minutes (A: 0.1 M Et_3NHOAc, pH 7.26; B: acetonitrile). The peak with the retention time of ~34.6 minutes was collected and lyophilized to dryness to give 1.43 nmoles of 3'-LEAE-5'-*CGCCTAAATG* (SEQ ID NO: 27) probe as determined from its UV absorbance at 260 nm. The probe was stored in 0.8 ml of 50 mM phosphate buffer, pH 6.0 containing 0.1% Bovine Serum Albumin (BSA) at –20 °C before use.

Oligonucleotides 5'-*TTCAGGTTCC* (SEQ ID NO: 28), 5'-*GACTTCTTCG* (SEQ ID NO: 29) and 5'-*AGATGGATGC* (SEQ ID NO: 30), all having an amino linker at the 3' end, are labeled with LEAE at the 3' end in the manner described above.

Example 5: Preparation of DMAE Labeled Detection Probes

Dimethyl acridinium esters (DMAE) are disclosed by Law et al. in U.S. Patent No. 4,745,181. These compounds emit light having an intensity maximum at the wavelength of 430 nm (λ_{max} = 430 nm).

In conjunction with the two color scheme described above and in Example 6 adapters of Example 1 for MPSS sequencing using the methods and sequences of the instant invention are encoded with nucleic aicd sequence for decoder binding. The DMAE is linked only to those decoder probes having complementary sequence to decoder binding sequence of those adapters with overhang sequences that incorporate either 8-oxo-dG or $\mathbf{X_2}$, such positions having a doubly degenerate base pairing set: {A, C}. The decoder binding sequences follow: 5'-CGCTTTGTAG (SEQ ID NO: 31); 5'-ATTCCTCCTC (SEQ ID NO: 32); 5'-GGCGATAACT (SEQ ID NO: 33); 5'-GCCAGTGTTA (SEQ ID NO: 34). The corresponding complementary decoder probe sequences (italics) are consequently:

5'-*CTACAAAGCG* (SEQ ID NO: 35); 5'-*GAGGAGGAAT* (SEQ ID NO: 36); 5'-*AGTTATCGCC* (SEQ ID NO: 37); 5'-*TAACACTGGC* (SEQ ID NO: 38).

The oligonucleotide, 5'-*CTACAAAGCG* (SEQ ID NO: 35)(8.5 nmoles), Is treated with triethylamine (536 umoles) for three hours at room temperature.

The DMAE-CO_2H was activated via mixed anhydride methods disclosed by Law et al. in U.S. Patent No. 5,622,825, as follows.

DMAE-CO_2H (2.5 mg, 5.36 μmoles) is dissolved in 1.5 ml of DMF and chilled in ice for several minutes. Triethylamine (6 μl, 42.9 μmoles) is added, followed by ethyl chloroformate (2.56 μl, 26.8 nmoles) and stirred, chilled, for half an hour. The reaction mixture is then dried with a rotary evaporator.

The residue is dissolved in DMF and the resulting activated DMAE- CO_2H (850 nmoles) added to the oligonucleotide, in a total volume of 300 μl of 1:1 DMF:H_2O. It is stirred at room temperature overnight.

The reaction mixture is passed through Sephadex G25 (fine) and eluted with water. The first peak was collected, concentrated by rotary evaporation and further purified by HPLC: (Column: Aquapore C8, RP-300, 4.6 mm x 25 cm (Rainin, Woburn, MA); Solvents: solvent A: 0.1 M Et_3NHOAc pH 7.2 – 7.4, solvent B: Acetonitrile; Gradient: (Linear) 8% to 20% B over 20 minutes, to 60% B over 20 minutes; Flowrate: 1 ml/minute; Detection λ: 254 nm). A product peak is collected and lyophilized to give 329 pmoles of the conjugate. The product is stored in 800 μl of 50 mM PO_4, pH 6.0, 0.1% BSA, at –20 °C prior to use.

Oligonucleotides 5'-*GAGGAGGAAT* (SEQ ID NO: 36); 5'-*AGTTATCGCC*

(SEQ ID NO: 37); 5'- *TAACACTGGC* (SEQ ID NO: 38) are labeled with DMAE at the 3' end in the manner described above.

Example 6: Two-Color MPSS with a Microbead Array

The MPSS ligation based sequencing method of Brenner et al. (2000), *supra*, is described in detail above. The sequences and methods of the instant invention are adapted to the MPSS method by employing the adapter sequences of Example 1 above and the two color decoder probe scheme for these adapters of Examples 4 and 5 above to the MPSS method.

The sequences are *in vitro* cloned onto the beads so that there are about 10^4-10^5 identical sequences per bead, and digestion is by the endonuclease *Bbv*I . Each cycle after the initial cleavage with *Dpn*II and fill in is summarized as follows: (i) ligation; (ii) detection by hybridization of decoder probes to decoder binding sites; (iii) *Bbv*I digestion. In the MPSS method without the instant invention, sixteen decoder binding sequences and decoder probes exist, which require sixteen cycles of decoder hybridizations to completely image the arrayed beads. As the methods and sequences of the instant invention reduce the number of adapters, decoder binding sequences and decoder probes to eight.

Use of decoder probes comprising only the sequences of the eight decoder probe sequences (SEQ ID NOS 27-30, 35-38) intrinsically labeled with ^{32}P as described in Example 2 above eight cycles may be used to completely image the signatures for each ligation/imaging/cleavage cycle.

Use of the two color chemiluminescent decoder probe labeling system of Examples 4 and 5 permits only imaging hybridization four subcycles per one ligation cycle.

Example 7: Two-Color MPSS Using Planar Spatial Substrate Surface Array

An array of the type described by Fodor et al in U.S. Patent No. 5,744,305 is constructed by, methods disclosed therein, preferably by presynthesizing oligonucleotides to be sequenced in parallel. *In situ* synthetic methods may be substituted with the caveat that the resulting array site regions will then not have as pure a population of the polymer intended for synthesis at the site. These consequently preferably *ex situ* made oligonucleotides are attached by now widely known

phosphoramidite chemistry adapted to photolithographic methods, e.g., by photolabile protecting groups used for masking. The array is constructed at a density of about 100 to 1,000,000 sites per cm^2, preferably at a density of about 1,000 to 100,000 sites per cm^2. All other aspects are as described in Example 6. The optional employment of the two colour visualization method permits streamlining the process so that only four decoder hybridization subcycles are required for complete imaging each ligation cycle.

Example 8: Classical SBH

The arrays of the type described in the preceding example can be adapted to perform the classical SBH. Instead of arraying analyte sequences, analysis of the types of sequences to be sequenced is performed by heuristic methods using bioinformatics and data specific to the species and type of DNA to be sequences. The SBH methods of Drmanac et al. (U.S. Patent No. 5,525,464) are described in more detail above. Analyte sequences are generated by PCR amplification with the ^{32}P labeling of Example 2 by use of the radioisotopically labeled dNTPs.

After analysis to determine the proper value of N, the length of the arrayed probes, and the proper length of the analyte fragments, the array is constructed. Instead of an array of all possible N-mers, a pair of N-mers each having a position with a unique partially overlapping doubly degenerate base pairing set is substituted for four possible N-mers having the standard nucleotides. Thus, for 8-mers, instead of four array sites having:

5'-NNNN(A)NNN;

5'-NNNN(T)NNN;

5'-NNNN(G)NNN;

5'-NNNN(C)NNN, two array sites are substituted.

The substituted two sites have the following probe sequences:

5'-NNNN(dP)NNN;

5'-NNNN(8-oxo-dG)NNN.

Alternatively the two sites substituted for the four are (X_1 and X_2 defined as above):

5'-NNNN(X_1)NNN;

5'-NNNN(X_2)NNN.

Or with adjustment of the density of polymers (NOT SITE DENSITY) to be twice as much for X_1 or X_2 compared to dP and 8-oxo-dG both the following are alternatively possible:

5'-NNNN(dP)NNN;

5'-NNNN(X_2)NNN; or

5'-NNNN(X_1)NNN;

5'-NNNN(8-oxo-dG)NNN.

Radioisotopically labeled analyte fragments may be visualized autoradiographically, or infrared photographic methods may be employed with unlabeled analyte fragments. Two analyte fragments could be simultaneously sequenced by two color methods employing the chemiluminescent labels of Examples 4 and 5.

Example 9: Allelic Analysis for Canavan Disease by PCR of Genomic DNA using Primer Sequences and Methods of the Invention

Canavan disease is an autosomal recessive disorder caused by aspartoacylase deficiency consequent acumulation of N-acetylaspartic acid in the brain. An A to C base change in nucleotide 854 of the open reading frame (ORF) of the human gene nucleic acid sequence, corresponding to nucleotide 1012 of the 1435 base pair long mRNA reverse transcribed cDNA , causing a missense mutation of amino acid 285 from glutamine (Glu) to alanine (Ala), has been shown to cause Canavan disease in the majority of alleles for the disease, with other mutations identified, as taught in U.S. Patent No. 5,697,635 to Matalon et al. Another mutation causing the disease is an ORF 693 mutation of C to A, resulting in the codon change TAC to TAA and a consequent termination instead of incorporation of Tyr 231. Yet another allele which has been identified is an ORF 914 position C to A change, causing the codon change of GCA to GAA for amino acid 305 in aspartoacylase, resulting in the missense mutation substituting a Glu (glutamic acid) for Ala 305.

An allelic analysis of genomic DNA by PCR, or of chromosome 17, easily separated by cytogenetic manipulative techniques, may be devised for either point mutation. The PCR amplification technique (see, for example, Mullis et al., U.S. Patent No. 4,683,202) and its requirements are widely appreciated. The mutation is detectable by dP and 8-oxo-dG probes comprising PCR primers of the invention, with the doubly degenerate base pairing nucleotides at the positions corresponding to, and pairing with,

values for hybridizing with different sequences, both for the same probe and compared with the pairing probe.

The sequence of the non-mutated sense strand of the human aspartoacylase gene beginning with ORF nucleotide 844 (1002 of the 1435 bp cDNA sequence) and ending in nucleotide 864 (1022 of the 1435 bp cDNA sequence) is 5'-TTTGTGAATG**A**GGCCGCATAT (SEQ ID NO: 39) (probed position bold underlined). This 21 base nucleotide sequence symmetric about ORF nucleotide 854 is complementary to 5'-ATATGCGGCC**T**CATTCACAAA (SEQ ID NO: 40).

The primers of the instant invention for allelic analysis are pairs of the complementary sequence, 5'-ATATGCGGCC**T**CATTCACAAA (SEQ ID NO: 40), with the probed position comprising doubly degenerate base pairing sets that partially overlap, e.g. 5'-ATATGCGGCC(ψ_i)CATTCACAAA (SEQ ID NO: 41), ψ_i, indicating either ψ_1 and ψ_2. Any of the partially overlapping ψ_1 and ψ_2 sets of Table 1 may be employed, ideally so that the mutation is amplified by both probes and the normal sequence is not amplified at all. The dP based primer, 5'-ATATGCGGCC**(dP)**CATTCACAAA (SEQ ID NO: 42), and 8-oxo-dG based primer, 5'-ATATGCGGCC**(8-oxo-dG)**CATTCACAAA (SEQ ID NO: 43), will amplify both the mutant and normal sequences of the A to C mutation of ORF base 854, while only the dG based primer will amplify the ORF 854 mutant (ORF 854 = C). Thus the afflicted homozygous mutated individual will exhibit amplification of both alleles by one probe, relative magnitude for simultaneous amplification 1 + 1 = 2, the carrier will exhibit amplification of the mutant allele by one primer and amplification of the non-mutated allele by both primers, relative magnitude 2 + 1 = 3, and the homozygous non-mutated individual will exhibit amplification of both alleles by both probes, relative magnitude 2 + 2 = 4. Thus, the three possibilities can be distinguished by quantifying the amplification product from simultaneous amplification using a combination of probes according to the invention. With $\mathbf{X_1}$ and $\mathbf{X_2}$ as defined above the, $\mathbf{X_1} = \Psi_1$ and $\mathbf{X_2} = \Psi_2$ based primer probes, 5'-ATATGCGGCC**(X_1)**CATTCACAAA (SEQ ID NO: 44), and 8-oxo-dG based primer, 5'-ATATGCGGCC**(X2)**CATTCACAAA (SEQ ID NO: 45) will function equivalently to the corresponding dP ($\mathbf{X_1}$) or 8-oxo-dG ($\mathbf{X_2}$) if their levels are doubled to effect the same effective number of primers for each base pairing of the degenerate set, and these may be substituted for one or both of the dP and 8-oxo-dG based primers. Note that, as defined, $\mathbf{X_1}$ based primers incorporate about equal amounts

of C and T at the probed position and X_2 based primers incorporate about equal amounts of G and T. Thus the 5'-ATATGCGGCC(**X_1**)CATTCACAAA (SEQ ID NO: 44) primer is actually a mixture of about equal amounts of:

5'-ATATGCGGCC**C**CATTCACAAA (SEQ ID NO: 46); and

5'-ATATGCGGCC**T**CATTCACAAA) (SEQ ID NO: 40)

The primer 5'-ATATGCGGCC(**X_2**)CATTCACAAA (SEQ ID NO: 45) is actually a mixture of about equal amounts of:

5'-ATATGCGGCC**G**CATTCACAAA (SEQ ID NO: 47); and

5'-ATATGCGGCC**T**CATTCACAAA (SEQ ID NO: 40).

Generally, more complicated potential allelic patterns, for example four possible nucleotides at the probed position, may be discerned by quantified amplification with the two primer probes separately, as described above. Except for *in utero* testing using **{dP or $X_1 = \psi_1$}** and **{8-oxo-dG or $X_2 = \psi_2$}** probes, which must identify the affected genotype, testing of adults for carrier screening in practice involves identifying reduced amplification product from quantitative simultaneous PCR with both primers. Known normal amplifications may be performed for calibration; the possibility of amplifying similar sequences from different genes is reduced by assaying only chromosome 17 pairs from the individual. Analogous primer pairs having the same partially overlapping doubly degenerate base pairing sets at the probed position can be employed for the other Canavan mutations described above for either simultaneous amplification of genomic DNA by both primers of the pair or separate amplification assays where the data is integrated after amplification. Individual chromosomes carrying the allele of interest can be separated to obtain more information, in some cases. In the Canavan context, separating the pair of chromosome 17 in the diploid somatic genome permits multiple primer pairs to be used to simultaneously screen the allele for several different amplification products that can be quantitatively distinguished for more detailed analysis, revealing some of the more rare mutations. Also, as will be appreciated by those skilled in the art is that these primers can also be used for screening based on cDNA derived from reverse transcription of expressed mRNA for the Canavan mutation. One important requirement for the operation of these primers with genomic DNA is that the DS primer sequence (e.g. a DS mutation centered sequence), may not be separated in the genomic DNA by untranslated intron sequence, which is spliced out in post-transcriptional processing. Thus, the probed position of the genomic DNA, for assays employing the

cDNA sequence, must not be so close to the splice junction that the sequence of the cDNA is not appropriate for the probe as some spliced out sequence is adjacent the probed position in the genomic DNA. The 854 ORF position mutation at 1012 of the 1435 base pair cDNA sequence of aspartoacylase is far from any intron exon junctions, being about in the middle of Exon 6 of the aspartoacylase gene which corresponds to positions 745 to 1270 of the 1435 base cDNA sequence (ORF 687-1112). For primer design for genomic DNA analysis of mutations near intron exon junctions, some of the mutation adjacent intron sequence must be known. The ORF 693 C to A mutation, for example, is close enough to the beginning of Exon 6 (ORF 687), that design of the primers of the invention for probing this position in genomic DNA is properly designed based in part upon the intron sequence preceding the beginning of Exon 6 (Intron 5 of the aspartoacylase gene), and primers for amplifying cDNA would be necessarily different than primers probing genomic sequence for this mutation (ORF 687 C to A).

A primer pair can be designed for the most common ORF 854 A to C mutation that causes Canavan disease, whereby the mutation sequence is amplified by both primers and the non-mutated sequence is not amplified at all. This would require doubly degenerate partially overlapping base pairing sets at the probed position that both include C as the common nucleotide in the base pairing set with A excluded from both base pairing sets: {C, T }; and {C, G}. Note that $\mathbf{Q_1}$, defined as about equal occupancy in the probed position of the bases A and G, has the first of the preceding base pairing sets, and $\mathbf{Q_2}$, about equal occupancy of bases G and C, will perform this function. The probe pair for the ORF 854 mutation is thus:

5'-ATATGCGGCC($\mathbf{Q_1}$)CATTCACAAA (SEQ ID NO: 48); and

5'-ATATGCGGCC($\mathbf{Q_2}$)CATTCACAAA) (SEQ ID NO: 49)

Again, the 5'-ATATGCGGCC($\mathbf{Q_1}$)CATTCACAAA (SEQ ID NO: 48) primer is actually a mixture of about equal amounts of:

5'-ATATGCGGCC**A**CATTCACAAA (SEQ ID NO: 50); and

5'-ATATGCGGCC**G**CATTCACAAA (SEQ ID NO: 47).

The primer 5'-ATATGCGGCC($\mathbf{Q_2}$)CATTCACAAA (SEQ ID NO: 49) is actually a mixture of about equal amounts of:

5'-ATATGCGGCC**G**CATTCACAAA (SEQ ID NO: 47); and

5'-ATATGCGGCC**C**CATTCACAAA (SEQ ID NO: 46).

mutant and non-mutant are desired to be detected. In the Canavan context, the array method may be preferable for *in utero* dignosis of the affected heterozygous mutants, and for screening the general population for carriers with the hope of discovering new single nucleotide polymorphisms at the probed positions, both pathologic (mutant) and non-pathologic.

Thus, an optimal probe pair for the 854 ORF mutation in such an array is:

5'-ATATGCGGCC($\underline{Q_1}$)CATTCACAAA (SEQ ID NO: 48); and

5'-ATATGCGGCC($\underline{Q_2}$)CATTCACAAA , (SEQ ID NO: 49) with Q_1 and Q_2 defined as in Example 9.

Again, the 5'-ATATGCGGCC($\underline{Q_1}$)CATTCACAAA (SEQ ID NO: 48)primer is actually a mixture of about equal amounts of:

5'-ATATGCGGCC<u>A</u>CATTCACAAA (SEQ ID NO: 50); and

5'-ATATGCGGCC<u>G</u>CATTCACAAA (SEQ ID NO: 47).

The primer 5'-ATATGCGGCC($\underline{Q_2}$)CATTCACAAA (SEQ ID NO: 49) is actually a mixture of about equal amounts of:

5'-ATATGCGGCC<u>G</u>CATTCACAAA (SEQ ID NO: 47); and

5'-ATATGCGGCC<u>C</u>CATTCACAAA (SEQ ID NO: 46). The other probe pairs are readily obtained analogously.